May 1963

A TWO-STAGE SAMPLING PROCEDURE FOR ESTIMATING A COMMON MEAN

Khursheed Alam[1]

Technical Report No. 25

University of Minnesota
Minneapolis, Minnesota

[1] This research was supported in part by the National Science Foundation under Grant Number G-19126.

Khursheed Alam

ABSTRACT

A TWO-STAGE SAMPLING PROCEDURE FOR ESTIMATING A COMMON MEAN

Consider $k \geq 2$ normal populations with a common mean μ and unknown variances $\sigma_1^2,\ldots,\sigma_k^2$, where $-\infty < \mu < \infty$ and $\sigma_1,\ldots,\sigma_k > 0$. Denote the populations by $\pi_1,\ldots,\pi_k$. It is desired to estimate μ by a two-stage sampling procedure with a fixed total number, say n, of available observations. The loss function is given by

$$L = \frac{n(\hat{\mu}-\mu)^2}{\min(\sigma_1^2,\ldots,\sigma_k^2)}$$

where $\hat{\mu}$ denotes the estimate of μ. The risk which is the expected value of L is denoted by $R_{\hat{\mu}}$.

Take a sample of equal size, m say, from each population in the first-stage experiment, where $n > mk$. Compute the sample variance of each population. Denote it by s_i^2 $(i=1,\ldots,k)$. In the second-stage take m_i observations from π_i $(i=1,\ldots,k)$ where $\sum_{i=1}^{k} m_i = n-mk$. Let $n_i = m_i+m$ denote the total number of observations taken from π_i and $\bar{x}_i$ denote the mean of the n_i observations. An estimate $\hat{\mu}$ of μ is given by

$$\hat{\mu} = \frac{\sum_{i=1}^{k} \frac{n_i\bar{x}_i}{s_i^2}}{\sum_{i=1}^{k} \frac{n_i}{s_i^2}} \qquad \text{--------- (1)}$$

where

$$n_i = n-m(k-1) \qquad \text{if } s_i^2 = \min(s_1^2,\ldots,s_k^2) = s^2(\text{say})$$

$$= m \qquad \text{if } s_i^2 > s^2$$

defines the sampling rule, which we denote by T.

For k=2 populations, Richter [1] obtained the following asymptotic results:

(i) As $n \to \infty$ $\sup_{\theta} R_{\hat{\mu}}(\theta) \to 1$ if and only if $m \to \infty$ and $\frac{m}{n} \to 0$, where $\theta = \frac{\sigma_1^2}{\sigma_2^2}$.

(ii) $m = (c_2 n)^{2/3} + 0(n^{1/3})$ minimizes $\sup_{\theta} R_{\hat{\mu}}(\theta)$, where $c_2 = .3399$.

(c_2 appears incorrectly in [1] but was corrected in a private communication).

(iii) $\min_{m} \sup_{\theta} R_{\hat{\mu}}(\theta) = 1 + c_2^{2/3} n^{-1/3} + 0(n^{-2/3})$.

It is shown in this paper that taking equal number of observations from each population in the first-stage is part of a minimax rule. It is also shown for k=2 populations that the rule T minimizes $\sup R_{\hat{\mu}}$ for any estimator $\hat{\mu}$ of the form (1) where $\sum_{i=1}^{k} n_i = n$ (the generalization of this result for $k > 2$ has been obtained but is omitted).

Let ψ denote the vector $(\theta_1,\ldots,\theta_k)$, where $\theta_i = \frac{\sigma^2}{\sigma_i^2}$ $(i=1,\ldots,k)$, $\sigma^2 = \min(\sigma_1^2,\ldots,\sigma_k^2)$, and $k \geq 2$. The following additional results are obtained in this paper:

(1) $R_{\hat{\mu}}(\psi)$ is a symmetric function of the components of ψ (hence we may assume that $\theta_k=1$).

(2) A necessary condition that $\lim_{n \to \infty} \sup R_{\hat{\mu}} = 1$ is that $m \to \infty$ and $\frac{m}{n} \to 0$ as $n \to \infty$.

(3) $R_{\hat{\mu}}(\psi)$ is maximized either at $\psi = (0,\ldots,0,1)$ at which its value is equal to $\frac{n}{n-m(k-1)}$ or at a point inside an interval I, where I is defined by $1-\theta_o \leq \theta_i \leq 1$, $i=1,\ldots,k$ and $0 < \theta_o = 0(m^{-1/2})$, or at both of these points.

(4) The minimax value of m is, therefore, a solution of the equation

$$\frac{m}{n-m(k-1)} = \sup_{\psi} R_{\hat{\mu}}(\psi)$$

(5) A solution of the above equation gives the unique asymptotic minimax value $m_k = \left(\frac{c_k n}{k-1} \right)^{2/3} + O(n^{1/3})$ where c_k is a constant. The values of c_k for $k=2(1)5$ are given in the appendix.

(6) The minimax risk is exactly equal to $\frac{n}{n-(k-1)m_k}$.

The supremum of the risk of an estimator of μ for a single stage sampling procedure, defined analogous to $\hat{\mu}$, is equal to k. A comparison of this value with the asymptotic value of Sup $R_{\hat{\mu}}$ which is equal to 1 shows the advantage of the two-stage sampling procedure over the single-stage sampling procedure.

When n-mk is large compared to m, it is shown that the risk of an estimator which is obtained by substituting the second sample variance for s^2 in $\hat{\mu}$, is smaller than $R_{\hat{\mu}}(\psi)$ for $\psi = (1,\ldots,1)$.

Consider $k \geq 2$ uniform populations with a common mean. With the same formulation of the problem as in the case of normal populations, it is shown that the minimax value of m is equal to $\left(\frac{nu_k}{2k} \right)^{1/2} + O(n^{1/4})$ where u_k is a constant; $u_2 = .5871$ approximately.

[1] Richter, E. (1960) "Two-stage experiments for estimating a common mean", Ann. Math. Statist. 31 pp. 1164-1173.

A TWO-STAGE SAMPLING PROCEDURE FOR ESTIMATING A COMMON MEAN

Khursheed Alam

1. Introduction. Let $\pi_1,\ldots,\pi_k$ be $k \geq 2$ normal populations with a common mean μ and variances $\sigma_1^2,\ldots,\sigma_k^2$ where $-\infty < \mu < \infty$, $\sigma_1,\ldots,\sigma_k > 0$. It is desired to estimate μ using a two-stage sampling procedure. The total number of available observations is fixed, say, n. In the absence of a priori information about the relative values of the variances of the populations we take a sample of equal size m, say, from each population in the first stage of the experiment. In the second stage we take m_i observations from π_i $(i=1,\ldots,k)$ where $\sum_{i=1}^{k} m_i = n-mk$. The values of the m_i's depend on the results of the first stage experiment. Using a suitable estimator of μ, which we denote by $\hat{\mu}$, we define the risk as

$$R_{\hat{\mu}} = \frac{nE(\hat{\mu}-\mu)^2}{\min(\sigma_1^2,\ldots,\sigma_k^2)} \qquad \text{-----} \ (1)$$

where E denotes expectation. It is required to determine the optimal values of $m,m_1,\ldots,m_k$. We shall investigate a minimax solution of the problem.

For an illustration, this problem may arise when there are available several devices for measuring a physical constant and the total number of measurements which can be taken is restricted by cost and time factors. Suppose that the cost of taking each measurement is equal to c, then the expected total loss involving n measurements may be expressed as

$$L = R_{\hat{\mu}}+cn$$

To minimize L we first minimize $R_{\hat{\mu}}$ for a given n.

For previous work done on this problem reference may be made to Richter [3] where it is shown that for $k=2$ the risk $R_{\hat{\mu}}$ which is bounded below by 1 converges uniformly to 1 as $n \to \infty$ provided that $m \to \infty$ and $\frac{m}{n} \to 0$; it is also shown that $m = (cn)^{2/3}$ is minimax value of m for large n, where $c = .34$ approximately.

Reference may also be made to Ghurye and Robbins [4] where it is required to estimate the difference between the means of two populations, using a two-stage sampling procedure. The cost of sampling is a known linear function of the number of observations taken from the two populations and is bounded above by a fixed number, say, A. It has been shown that the ratio of the variance of a suitable estimator to the variance of the optimal estimator which requires that the number of observations be distributed between the two populations in certain proportion depending on the true variances of the two populations, tends to unity as $A \to \infty$.

The estimator $\hat{\mu}$ is defined as follows. We denote by s_i^2 the estimate (sample variance) of the variance of π_i $(i=1,\ldots,k)$ obtained from the first stage experiment. Thus, each s_i^2 is based on m observations. Let $n_i = m_i+m$ denote the total number of observations taken from π_i and $\bar{x}_i$ denote the mean of those n_i observations. Then

$$\hat{\mu} = \frac{\sum_{i=1}^{k} \frac{n_i \bar{x}_i}{s_i^2}}{\sum_{i=1}^{k} \frac{n_i}{s_i^2}} \qquad \text{- - - - - (2)}$$

where

$$n_i = n-m(k-1) \qquad \text{if } s_i^2=s^2=\min(s_1^2,\ldots,s_k^2)$$

$$= m \qquad \text{if } s_i^2 > s^2$$

$$\sum_{i=1}^{k} n_i = n$$

If $s_i^2=s^2$ for more than one value of i, an event of probability measure zero, we take one of them to be the smallest among themselves by selection through any random procedure.

From the definition of $\hat{\mu}$ we note that the sampling procedure requires that all observations should be taken in the second-stage experiment from the population which yelds the smallest sample variance in the first-stage experiment. It is shown in this paper, for k=2 populations, that this procedure is minimax. This result can be generalized for $k > 2$. The m_i's being, thus, determined by the sampling rule, the remaining problem reduces to the choice of an optimal value of m.

Let ψ denote the vector $(\theta_1,\dots,\theta_k)$ where $\theta_i = \frac{\sigma^2}{\sigma_i^2}$ (i=1,...,k) and $\sigma=\min(\sigma_1,\dots,\sigma_k)$. It is shown below that $R_{\hat{\mu}}(\psi)$ is a symmetric function of the components of ψ. Without loss of generality, therefore, we can take $\theta_k=1$. $R_{\hat{\mu}}(\psi)$ is bounded below by 1. A necessary condition that $\lim_{n\to\infty} \sup R_{\hat{\mu}} = 1$ is that

$$m \to \infty \quad \text{and} \quad \frac{m}{n} \to 0 \quad \text{as} \quad n \to \infty.$$

$R_{\hat{\mu}}(\psi)$ is maximized either at $\psi=(0,\dots,0,1)$ at which its value is equal to $\frac{n}{n-mk+m}$ or inside the interval I, defined as

$$1-\theta_o \le \theta_i \le 1 \qquad i=1,\dots,k$$

where $0 < \theta_o = O(\frac{1}{m^{\frac{1}{2}}})$. Hence the minimax value of m is a solution of the equation

$$\frac{n}{n-mk+m} = \sup_{\psi \in I} R_{\hat{\mu}}(\psi)$$

An asymptotic solution of the above equation gives the value of m, which we denote by m_k, as

$$m_k = \left(\frac{c_k n}{k-1}\right)^{2/3}$$

where c_k (k denoting the number of populations) is a constant. $c_2 = .3399$ approximately. The value of the minimax risk is equal to $\frac{n}{n-m_k k+m}$. These results are generalizations for $k \geqq 2$ populations of similar results obtained by Richter for the case of two populations, to which a reference has been made above.

It will be observed from (2) that the quantities s_i^2 $(i=1,\ldots,k)$ involved in the expression for $\hat{\mu}$ are used as estimates of the variances of the k populations. These estimates are obtained from the first-stage experiment. We might use the second sample to improve the estimate of the smallest of the population variances. It is shown that when n-mk is large compared to m the risk of an estimator $\hat{\delta}$ which is obtained by substituting the second sample variance for the smallest of $s_1^2,\ldots,s_k^2$ in the expression for $\hat{\mu}$ has a smaller risk than $R_{\hat{\mu}}$.

In single-stage sampling procedure, with a fixed total of n available observations, we take a sample of size $\frac{n}{k}$ from each population, assuming n is a multiple of k. It is shown that the supremum of the risk of an estimator, which is defined analogous to $\hat{\mu}$, is equal to k. A comparison of this value with the asymptotic value of Sup $R_{\hat{\mu}}$ which is equal to 1 shows the advantage of the two-stage sampling procedure over the single-stage sampling procedure, under the criterion of minimaxity.

In the last section of this paper we consider $k \geqq 2$ uniform populations with a common mean. The formulation of the problem is the same as in the case of normal populations. The analysis is similar. Only the main results are, therefore, given without showing the derivation.

We have k rectangular populations $\pi_1,\ldots,\pi_k$ with a common mean equal to μ. We shall denote the range of π_i by a_i $(i=1,\ldots,k)$. The quantities a_i and μ are unknown. It is required to estimate μ by a two-stage sampling procedure. The total number of available observations is fixed equal to n, say. We take $m \geqq 2$ observations from each population and take $n-mk$ observations in the second-stage experiment from the population corresponding to the smallest range of the first sample. The estimate of μ is given by

$$\hat{\mu} = \frac{x_{(1)}+x_{(n-mk)}}{2}$$

where $x_{(1)}$, $x_{(n-mk)}$ are the extreme values of the observations in the second sample. The risk of $\hat{\mu}$, which we shall denote by R, is defined as

$$R = \frac{2(n+1)(n+2)}{a^2} E(\hat{\mu}-\mu)^2$$

where $a = \min(a_1,\ldots,a_k)$.

The asymptotic minimax value of m, which we denote by m_k, is given as

$$m_k = \left(\frac{nu_k}{2k}\right)^{\frac{1}{2}} + O(n^{\frac{1}{4}})$$

and the value of the minimax risk is equal to

$$\frac{(n+1)(n+2)(1+\frac{u_k}{m_k})}{(n-m_k k+1)(n-m_k k+2)}$$

where u_k is a constant. $u_2 = .5871$ approximately. Thus, the minimax value of m is of order $n^{\frac{1}{2}}$ as against $n^{2/3}$ for the normal populations. Also, the value of the minimax risk is equal to $1+O(\frac{1}{n^{\frac{1}{2}}})$ as against $1+O(\frac{1}{n^{1/3}})$ for the normal populations.

In section 2 we show symmetry of the problem under a class of transformations of the sample space and the relation of $\hat{\mu}$ to an optimal invariant estimator of μ when the variances of the populations are known. In section 3 it is shown, for the case of two populations, that the procedure of taking all observations in the second sample from the population corresponding to the smallest variance of the first sample is minimax. In section 4 we give a lower and an upper bound on the value of $R_{\hat{\mu}}$. In section 5 we show that the minimax value of m is of order $n^{2/3}$. In section 6 it is shown that the supremum of $R_{\hat{\mu}}$ is attained at one of two points. The minimax value of m and the value of the minimax risk are derived in section 7. In section 8 we show that the supremum of the risk of an estimator of μ for a single-stage sampling procedure is equal to k. In section 9 it is shown that the risk is reduced by using the second sample variance for estimating the smallest of the population variances. Some results with respect to uniform populations are given in section 10.

2. Symmetry Of The Problem. The statistical problem of the estimation of the common mean of the k populations remains invariant under a group G of linear transformations of the sample space. G is defined as follows. Let $x_1,\ldots,x_n$ be a set of n observations from the k populations. Then for $g \in G$

$$g(x_1,\ldots,x_n) = (ax_1+b,\ldots,ax_n+b)$$

for some $a \neq 0$ $-\infty < b < \infty$. For the corresponding transformation on the parameter space we have

$$\bar{g}(\mu,\sigma_1^2,\ldots,\sigma_k^2) = (a\mu+b,\ a^2\sigma_1^2,\ldots,a^2\sigma_k^2)$$

Clearly, the family of the underlying distributions and the structure of the loss function of the problem remains invariant under this class of transformations. An estimator δ is said to be invariant under G if it satisfies the following relation

$$\delta(ax_1+b,\ldots,ax_n+b) = a\delta(x_1,\ldots,x_n)+b \qquad \text{- - - - - (3)}$$

for all $x_1,\ldots,x_m$, $a \neq 0$, $-\infty < b < \infty$.

The problem is also invariant under the group, call it G', of all permutations of the labels of the k populations. Under the principle of invariance we may restrict the choice of an estimator of μ to the class of estimators invariant under G and G'. It is easily seen that $\hat{\mu}$ defined in (2) is a member of this class. By the invariance under G it follows ([1] page 226) that $R_{\hat{\mu}}$ is a function of $\theta_1,\ldots,\theta_k$. Further, by the invariance under G' it follows that

<u>Theorem 1</u> $R_{\hat{\mu}}(\psi)$ is a symmetric function of the components of ψ. For the particular case when k=2, denoting $\frac{\sigma_1^2}{\sigma_2^2}$ by θ, we have

$$R_{\hat{\mu}}(\theta) = R_{\hat{\mu}}(\tfrac{1}{\theta})$$

It is easy to see that $\hat{\mu}$ is also an unbiased estimator of μ. We shall now show its relation to an optimal invariant estimator of μ when the variances of the populations are known, and the experiment is given, that is the number of observations taken from each population is fixed.

If y is a vector random variable with k components having a density function which is known except for a location parameter μ, that is, the density is given by

$$p(y|\mu) = q(y-\mu\epsilon)$$

where q is a known function and $\epsilon=(1,\ldots,1)$. Then it is known ([1] page 310) that the optimal invariant (with respect to translation) estimate of μ with squared error as loss function is

$$y_1 - E(y_1-\mu|y_2-y_1,\ldots,y_k-y_1) \qquad ----- (4)$$

Thus, supposing that $\sigma_1,\ldots,\sigma_k$ are known and that the n_i's are fixed, we get the optimal invariant (with respect to translation) estimator of μ as

$$\frac{\sum_{i=1}^{k} \frac{n_i \bar{x}_i}{\sigma_i^2}}{\sum_{i=1}^{k} \frac{n_i}{\sigma_i^2}} \quad \text{- - - - - (5)}$$

The quantity $\hat{\mu}$ given in (2) is obtained from (5) by substituting s_i^2 for the unknown parameter σ_i^2 $(i=1,\ldots,k)$. Observe that s_i^2 converges strongly to σ_i^2 as m becomes large when σ_i is bounded.

We observe that, given the sampling rule, a sufficient statistic for the estimation of μ is (u, V) where $u=\bar{x}_1,\ldots,\bar{x}_k$, $s_1^2,\ldots,s_k^2$ and V is the sample variance computed from n-m(k-1) observations drawn from the population for which $s_i^2=s^2$. However, the distribution of the random variable V is difficult to handle mathematically. For the sake of simplicity, therefore, we consider only the class of estimators of μ which are functions of u. The estimator $\hat{\mu}$ is a member of this class.

3. Sampling Rule. Clearly, the choice of a sampling rule is related to the method of estimation followed after the sample is obtained. In this section we study the sampling rule as defined following (2) which we shall denote by T. From (2) we obtain a class of estimators of μ by varying the sampling rule such that the m_i's are positive integral valued functions of $s_1,\ldots,s_k$. By symmetry we further restrict the m_i's to depend on $s_1,\ldots,s_k$ only through the ratios of these quantities. Let D be the class of all estimators of μ thus obtained. It is shown below for the case of k=2 populations that $\hat{\mu}$ as defined in (2) is minimax in D. This result can be generalized for $k > 2$ populations. It is also shown that $\hat{\mu}$ is asymptotically optimal in D.

If it is known a priori which of the k populations has the smallest variance then it is optimal to take all observations from that population, say, π_1. In this case the estimator of μ is $\bar{x}_1$, the mean of n observations from π_1. The corresponding value of the risk is equal to 1 which, as we shall see in the next section, is a lower bound of $R_{\hat{\mu}}$.

In the absence of a priori information about the variances of the populations, we take equal number of observations from each population in the first stage of the experiment. Equality of the sample size in the first stage is required by consideration of the symmetry of the problem under permutations of the labels of the populations. As all permutations of the labels of the populations make a finite group of transformations, it follows ([1] page 226) that the procedure of taking equal number of observations in the first stage of the experiment is minimax.

Consider the case k=2. From (2) we easily compute

$$R_{\hat{\mu}} = E_S \frac{n(n_1\theta+n_2S^2)}{\min(1,\theta)(n_1+n_2S)^2}$$

$$= E_F \frac{n\theta(n_1+n_2F^2\theta)}{\min(1,\theta)(n_1+n_2F\theta)^2} \qquad \text{- - - - - (6)}$$

where $F=F_{m-1,m-1}$ in Fisher's ratio of two independent chi-square variables, each with m-1 degrees of freedom, $\theta=\frac{\sigma_1^2}{\sigma_2^2}$ and $S=\frac{s_1^2}{s_2^2}$. It was shown in the previous section that for any θ

$$R(\theta) = R(\frac{1}{\theta})$$

Let $(\theta_o, \frac{1}{\theta_o})$ be two values of θ at which $R_{\hat{\mu}}(\theta)$ is maximized, where $\theta_o \geq 1$. Assume an a priori distribution for θ such that

$$P\{\theta=\theta_o\} = P\{\theta=\frac{1}{\theta_o}\} = \tfrac{1}{2}$$

For the posterior distribution of θ, given S, we have

$$P\{\theta=\theta_o|S\} = A(S+\theta_o)^{-m+1}$$

$$P\{\theta=\frac{1}{\theta_o}|S\} = A(1+S\theta_o)^{-m+1}$$

where $A = \{(S+\theta_o)^{-m+1}+(1+S\theta_o)^{-m+1}\}^{-1}$. For the posterior risk of $\hat{\mu}$ given S we have then

$$R_{\hat{\mu}}(S) = \frac{nA}{(n-m+mS)^2}\{((n-m)\theta_o+mS^2)(S+\theta_o)^{-m+1}+(n-m+mS^2\theta_o)(1+S\theta_o)^{-m+1}\} \text{ for } S \leq 1$$

$$= \frac{nA}{(m+(n-m)S)^2}\{(m\theta_o+(n-m)S^2)(S+\theta_o)^{-m+1}+(m+(n-m)S^2\theta_o)(1+S\theta_o)^{-m+1}\} \text{ for } S > 1$$

Similarly for the posterior risk of $\hat{\delta}$, any estimator in D, we have, assuming k=2

$$R_{\hat{\delta}}(S) = E_\theta \frac{n(n_1\theta+n_2S^2)}{\min(1,\ \theta)(n_1+n_2S)^2}$$

$$= \frac{nA}{(n_1+n_2S)^2}\{(n_1\theta_o+n_2S^2)(S+\theta_o)^{-m+1}+(n_1+n_2S^2\theta_o)(1+S\theta_o)^{-m+1}\}$$

Consider the case $S \leq 1$. Then

$$R_{\hat{\delta}}(S)-R_{\hat{\mu}}(S) = nA[\{\frac{n_1\theta_o+n_2S^2}{(n_1+n_2S)^2} - \frac{(n-m)\theta_o+mS^2}{(n-m+mS)^2}\}(S+\theta_o)^{-m+1}$$

$$+ \{\frac{n_1+n_2S^2\theta_o}{(n_1+n_2S)^2} - \frac{n-m+mS^2\theta_o}{(n-m+mS)^2}\}(1+S\theta_o)^{-m+1}]$$

It can be shown that the quantity inside the second brace on the right is positive and that $S+\theta_o \geq 1+S\theta_o$. Hence for $m \geq 2$

$$R_{\hat{\delta}}(S)-R_{\hat{\mu}}(S) \geq nA(S+\theta_o)^{-m+1}\{\frac{n_1\theta_o+n_2S^2+n_1+n_2S^2\theta_o}{(n_1+n_2S)^2} - \frac{(n-m)\theta_o+mS^2+(n-m)+mS^2\theta_o}{(n-m+mS)^2}\}$$

$$= nA(S+\theta_o)^{-m+1}(1+\theta_o)\{\frac{n_1+n_2S^2}{(n_1+n_2S)^2} - \frac{n-m+mS^2}{(n-m+mS)^2}\}$$

$$= \frac{nA(S+\theta_o)^{-m+1}(1+\theta_o)(S-1)^2\{n_1(n-m)+mn_2S^2\}(n_2-m)}{(n_1+n_2S)^2(n-m+mS)^2}$$

$$\geq 0$$

Similarly it can be shown that for $S > 1$

$$R_{\hat{\delta}}(S)-R_{\hat{\mu}}(S) \geq 0$$

We conclude that $\hat{\mu}$ is a Bayes estimate with respect to the a priori distribution (two point distribution) of θ which is stated above. Also, it is essentially unique. We now make use of the following lemma, due to Lehmann ([7] page 4-20), which we state without proof.

Lemma: Let x have a distribution $p_\theta(x)$, $\theta\epsilon\Omega$. Suppose there is a distribution λ over Ω and a set $\omega\epsilon\Omega$ such that

$$R_{\delta_\lambda}(\theta) = \sup_{\theta\epsilon\Omega} R_{\delta_\lambda}(\theta) \qquad \text{for all } \theta\epsilon\omega$$

where δ_λ is the Bayes procedure with respect to λ and $\lambda(\omega)=1$ then δ_λ is minimax. Further if δ_λ is unique Bayes then δ_λ is unique minimax and hence also admissible. Here R_{δ_λ} denotes the risk of δ_λ.

From this lemma we obtain

Theorem 2 $\hat{\mu}$ is minimax and admissible in the class D.

We shall now show that $\hat{\mu}$ is asymptotically optimal in D. From (2) we have

$$R_{\hat{\mu}} = \frac{n}{\sigma^2} E \frac{\sum_{i=1}^{k} \frac{n_i\sigma_i^2}{s_i^4}}{\left(\sum_{i=1}^{k} \frac{n_i\sigma_i^2}{s_i^2}\right)^2}$$

$$= n\,E\,\frac{\sum_{i=1}^{k} \frac{n_i \theta_i}{x_i^2}}{\left(\sum_{i=1}^{k} \frac{n_i \theta_i}{x_i}\right)^2} \qquad ----- (7)$$

where $\{x_i;\ i=1,\ldots,k\}$ are k mutually independent chi-square variables, each with m-1 degrees of freedom

$$n_i = n-mk+m \quad \text{if } \frac{x_i}{\theta_i} = \min\left(\frac{x_1}{\theta_1},\ldots,\frac{x_k}{\theta_k}\right)$$

$$= m \qquad \text{otherwise}$$

$$\sum_{i=1}^{k} n_i = n$$

and $\theta_i = \frac{\sigma^2}{\sigma_i^2}$ $(i=1,\ldots,k)$. Notice that for each i

$$0 \leq \theta_i \leq 1$$

and that one of the θ_i's is identically equal to 1. As we have seen earlier that $R_{\hat{\mu}}(\theta_1,\ldots,\theta_k)$ is a symmetric function of $\theta_1,\ldots,\theta_k$ we shall henceforth assume, without any loss of generality, that $\theta_k=1$.

When m is large we obtain form (7) putting x_i=m-1, its expected value, for each i

$$R_{\hat{\mu}} \approx n\,E\,\frac{1}{\sum_{i=1}^{k} n_i \theta_i} \qquad ----- (8)$$

$$\approx \frac{n}{m \sum_{i=1}^{k-1} \theta_i + (n-mk+m)} \qquad ----- (9)$$

The quantity on the right of (8) represents the asymptotic value of $R_{\hat{\delta}}$ for any $\hat{\delta} \in D$. From (8) and (9) it is easy to see that

$$R_{\hat{\delta}} \geq R_{\hat{\mu}}$$

when m is sufficiently large.

On the basis of the results of the first-stage experiment a minimal sufficient statistic is $w = \bar{y}_1,\ldots,\bar{y}_k,\ s_1^2,\ldots,s_k^2$ where $\bar{y}_1,\ldots,\bar{y}_k$ denote the sample means. By sufficiency, the m_i's, denoting the second-stage sample size, are reduced to functions of w. Under invariance with respect to the group of linear transformations, discussed in the previous section, we require that for each i

$$m_i(a\bar{y}_1+b,\ldots,a\bar{y}_k+b,\ s_1^2,\ldots,s_k^2) = m_i(\bar{y}_1,\ldots,\bar{y}_k,\ s_1^2,\ldots,s_k^2)$$

$$a = \pm 1, \quad -\infty < b < \infty$$

A maximal invariant of the arguments under translation and change of sign is given by

$$(\bar{y}_1-\bar{y}_k)^2,\ldots,(\bar{y}_{k-1}-\bar{y}_k)^2, \quad s_1^2,\ldots,s_k^2$$

However,

$$m(\bar{y}_i-\bar{y}_k)^2 \overset{d}{\approx} (\sigma_i^2+\sigma_k^2)\chi_1^2 \qquad i=1,\ldots,k-1$$

$$m^2 s_i^2 \overset{d}{\approx} \sigma_i^2\chi_{m-1}^2 \qquad i=1,\ldots,k$$

where $\overset{d}{\approx}$ means 'is distributed as' and χ_p^2 denotes a chi-square random variable with p degrees of freedom. Thus, the quantities $(\bar{y}_1-\bar{y}_k)^2,\ldots,(\bar{y}_{k-1}-\bar{y}_k)^2$ are relatively un-informative with regard to the estimation of the parameters $\sigma_1^2,\ldots,\sigma_k^2$ compared to the quantities $s_1^2,\ldots,s_k^2$ in the following sense. If an experiment results in a random variable x which is distributed as $u\chi_1^2$ and another independent experiment results in a random variable y which is distributed as $u\chi_p^2$ where u is a positive valued unknown parameter, then for the estimation of u, y is p times as informative as x because y is equivalent to p independent values of x. That is to say, the second experiment is equivalent to p replications of the first experiment. If, therefore, m is large we can suppose, without any appreciable loss of precision, that the m_i's are functions of $s_1^2,\ldots,s_k^2$. Invariance under scalar transformation further reduces the m_i's to functions of the ratios of $s_1^2,\ldots,s_k^2$. The above discussion is given for an appreciation of the choice of the sampling rule T.

4. Bounds On the Risk of $\hat{\mu}$. In this section we obtain bounds on the risk of $\hat{\mu}$. The results will be used in later sections to obtain an asymptotic minimax value of m. We shall denote the quantity on the right of (8) by R_1. It is easy to see that R_1 is minimized at $\theta_1 = \ldots = \theta_k = 1$ and that the minimum value is equal to 1. By Schwarz inequality we have

$$\left(\sum_{i=1}^{k} \frac{n_i\theta_i}{x_i^2} \right) \left(\sum_{i=1}^{k} n_i\theta_i \right) \geq \left(\sum_{i=1}^{k} \frac{n_i\theta_i}{x_i} \right)^2 \qquad ----- (10)$$

for any set of non-negative values of the x_i's and the θ_i's. Hence

$$R_{\hat{\mu}} \geq R_1 \geq 1 \qquad ----- (11)$$

From (7) we have

$$\frac{1}{n} R_{\hat{\mu}} = \sum_{i=1}^{k} E\left\{ \frac{\sum_{\alpha=1}^{k} \frac{n_\alpha\theta_\alpha}{x_\alpha^2}}{\left(\sum_{\alpha=1}^{k} \frac{n_\alpha\theta_\alpha}{x_\alpha} \right)^2} \middle| n_i=n-mk+m \right\} p\{n_i=n-mk+m\}$$

$$\leq \sum_{i=1}^{k} E\left\{ \frac{m \sum_{\alpha\neq i} \frac{\theta_\alpha x_i^2}{\theta_i x_\alpha^2} + n-mk+m}{(n-mk+m)^2\theta_i} \middle| n_i=n-mk+m \right\} p\{n_i=n-mk+m\}$$

$$\leq \sum_{i=1}^{k} E\{ \frac{m \sum_{\alpha \neq i} \frac{x_i}{x_\alpha} + n-mk+m}{(n-mk+m)^2 \theta_i} | n_i=n-mk+m\} p\{n_i=n-mk+m\}$$

$$\leq \sum_{i=1}^{k} E\{ \frac{m \sum_{\alpha \neq i} \frac{x_i}{x_\alpha} + n-mk+m}{(n-mk+m)^2 \theta_i} \} p\{n_i=n-mk+m\}$$

In deriving the last inequality we have used the fact that

$$E\{u|u \leq v\} \leq E(u)$$

where u and v are any two random variables. We have then

$$\frac{(n-mk+m)^2}{n} R_{\hat{\mu}} \leq \sum_{i=1}^{k} \{ \frac{m(k-1)(m-1)}{(m-3)} + (n-mk+m)\} \frac{p\{n_i=n-mk+m\}}{\theta_i}$$

$$= \{ \frac{m(k-1)(m-1)}{m-3} + (n-mk+m)\} [1 + \sum_{i=1}^{k-1} \frac{1-\theta_i}{\theta_i} p\{n_i=n-mk+m\}] \quad - - - - (12)$$

Now

$$p\{n_i=n-mk+m\} = \int_0^\infty \prod_{j \neq i} \{1-v(\frac{\theta_j}{\theta_i} x)\} dv(x)$$

$$\leq \int_0^\infty \{1-v(\frac{x}{\theta_i})\} dv(x) \qquad \text{for } i \neq k$$

$$= p\{F > \frac{1}{\theta_i} \} \qquad - - - - - (13)$$

where $v(x)$ is cumulative distribution function of a chi-square variable with m-1 degrees of freedom and $F = F_{m-1,m-1}$ is the ratio of two independent chi-square variables, each with m-1 degrees of freedom. Hence from (12) we have

$$\frac{(n-mk+m)^2}{n} R_{\hat{\mu}}$$

$$\leq \left\{ \frac{m(k-1)(m-1)}{m-3} + (n-mk+m) \right\} \left[1 + \sum_{i=1}^{k-1} \frac{1-\theta_i}{\theta_i} p\{F > \frac{1}{\theta_i}\} \right] \quad \text{- - - - (14)}$$

Let $A(\theta) = (\theta-1)p\{F > \theta\}$. Differentiating with respect to θ we have

$$\frac{\partial^2 A}{\partial \theta^2} = \frac{\theta^{\frac{m-5}{2}} (1+\theta)^{-m} (m-3)}{2B(\frac{m-1}{2}, \frac{m-1}{2})} \left(\theta^2 - \frac{2(m+1)\theta}{m-3} + 1\right) \quad \text{- - - - - - (15)}$$

where $B(\frac{m-1}{2}, \frac{m-1}{2}) = \dfrac{\Gamma(\frac{m-1}{2})\,\Gamma(\frac{m-1}{2})}{\Gamma(m-1)}$ is the complete beta function.

Note that the density function of F is given by

$$g(F) = \frac{F^{\frac{m-3}{2}} (1+F)^{-m+1}}{B(\frac{m-1}{2}, \frac{m-1}{2})} \qquad F > 0$$

From (15) we find that for $\theta \geq 1$

$$\frac{\partial^2 A(\theta)}{\partial \theta^2} < 0 \qquad 1 \leq \theta < \theta_1$$

$$= 0 \qquad \theta = \theta_1$$

$$> 0 \qquad \theta > \theta_1$$

where $\theta_1 = \dfrac{(m+1) + 2\sqrt{2(m-1)}}{m-3}$ is the larger of the two roots of the equation

$$\theta^2 - \frac{2(m+1)\theta}{m-3} + 1 = 0$$

It follows that in the range $\theta \geq 1$, $A(\theta)$ is maximized at $\theta = \theta_o$ where

$$1 < \theta_o < \theta_1 \quad \text{- - - - - (16)}$$

Hence for $\theta \geqq 1$

$$A(\theta) \leq \tfrac{1}{2}(\theta_1 - 1) = \frac{2}{m-3} + \frac{\sqrt{2(m-1)}}{m-3} \leq \frac{2}{m-3} + \frac{2}{\sqrt{m-3}} \quad \text{for } m \geqq 5$$

From (14) it follows that for $m \geqq 5$

$$\frac{(n-mk+m)^2}{n} R_{\hat{\mu}}$$

$$\leq \left\{ \frac{(k-1)m(m-1)}{m-3} + (n-mk+m)\right\}\left\{1+2(k-1)\left(\frac{1}{m-3} + \frac{1}{\sqrt{m-3}} \right)\right\} \quad ----- (17)$$

From (8) and (9) we have

$$R_{\hat{\mu}} - R_{\hat{\mu}_1} = nE \frac{\sum\limits_{\substack{i,j=1,\dots,k \\ j>i}} n_i n_j \theta_i \theta_j \left(\frac{x_j - x_i}{x_j x_i} \right)^2}{\left(\sum\limits_{i=1}^{k} \frac{n_i \theta_i}{x_i} \right)^2 \left(\sum\limits_{i=1}^{k} n_i \theta_i \right)}$$

$$\leq \frac{n}{(n-mk+m)^2} E \frac{\sum\limits_{i,j=1,\dots,k \; j>i} n_i n_j \theta_i \theta_j \left(\frac{x_j - x_i}{x_j x_i} \right)^2 x_k^2}{\sum\limits_{i=1}^{k} n_i \theta_i}$$

now

$$\frac{n_i n_j \theta_i \theta_j}{\sum\limits_{i=1}^{k} n_i \theta_i} \leq m \qquad \text{for all } i,j$$

Hence

$$R_{\hat{\mu}} - R_1 \leq \frac{mn}{(n-mk+m)^2} E \sum_{\substack{i,j=1,\dots,k \\ j>i}} \left(\frac{x_j - x_i}{x_j x_i} \right)^2 x_k^2$$

$$= \frac{mn}{(n-mk+m)^2} E\left\{ \sum_{j=2}^{k} \left(\frac{x_j - x_i}{x_j} \right)^2 + E(x_1^2)\, E \sum_{\substack{j>i \\ i \neq 1}} \left(\frac{x_j - x_i}{x_j x_i} \right)^2 \right\}$$

$$= \frac{mn}{(n-mk+m)^2} \quad \{(k-1)(1 - \frac{2(m-1)}{m-3} + \frac{m^2-1}{(m-3)(m-5)})$$

$$+ (k-1)(k-2)(\frac{1}{(m-3)(m-5)} - \frac{1}{(m-3)^2})\}$$

$$= \frac{mn(k-1)}{(n-mk+m)^2} \{ \frac{4(m-4)}{(m-3)(m-5)} + \frac{2(k-2)(m^2-1)}{(m-3)^2(m-5)} \} = 0(\frac{1}{n}) \quad \text{------} (18)$$

5. Asymptotic Minimax value of m. From the expression for R_1 given in (8), putting $\theta_1 = \theta_2 = \ldots = \theta_{k-1} = \theta$, we get

$$R_1(\theta,\ldots,\theta,1) = \frac{n}{n-mk+m+m(k-1)\theta} \int_0^\infty \{1-v(x\theta)\}^{k-1} dv(x)$$

$$+ \frac{n}{m+(n-m)\theta} [1 - \int_0^\infty \{1-v(x\theta)\}^{k-1} dv(x)] \quad \text{----} (19)$$

For k=2 we get

$$R_1(\theta, 1) = \frac{n}{n-m+m\theta} \int_0^\infty \{1-v(x\theta)\} dv(x)$$

$$+ \frac{n}{m+(n-m)\theta} [1 - \int_0^\infty \{1-v(x\theta)\} dv(x)] \quad \text{-----} (20)$$

From (19) and (20) it is easy to show that for $k \geq 2$

$$R_1(\theta,\ldots,\theta,1) \geq R_1(\theta, 1) \quad \text{------} (21)$$

We can write (20) as

$$R_1(\theta, 1) = \frac{n}{n-m+m\theta} p\{F > \theta\} + \frac{n}{m+(n-m)\theta} p\{F \leq \theta\}$$

$$\geq \{1 + \frac{m(1-\theta)}{n} \}p\{F > \theta\} + \{1 + \frac{(n-m)(1-\theta)}{n} \}p\{F \leq \theta\}$$

$$\geq 1 + \frac{m(1-\theta)}{2n} + \frac{(n-m)(1-\theta)}{n} \, p\{F \leq \theta\} \qquad \text{- - - - - (22)}$$

As in (16) we find that for $\theta \leq 1$ the maximum value of $(1-\theta)p\{F \leq \theta\}$ is obtained at $\theta = 1 - \frac{2t}{\sqrt{m}}$ where $t = O(1)$. From Edgeworth's approximation of distribution function we find that

$$p\{F \leq 1 - \frac{2t}{\sqrt{m}}\} = \Phi(-t) + O(\frac{1}{\sqrt{m}})$$

Hence from (22) we have

$$R_1(\theta, 1) \geq 1 + \frac{m(1-\theta)}{2n} + \frac{2t(n-m)}{nm^{\frac{1}{2}}} \{\Phi(-t) + O(\frac{1}{m^{\frac{1}{2}}})\}$$

Therefore, from (11) and (21) we get the inequality

$$R_1(\theta,\ldots,\theta,1) \geq 1 + \frac{m(1-\theta)}{2n} + \frac{2t(n-m)}{nm^{\frac{1}{2}}} \{\Phi(-t) + O(\frac{1}{m^{\frac{1}{2}}})\} \qquad \text{- - - (23)}$$

It follows from the above inequality that if $m = o(n^{2/3})$, or if $n = o(m^{3/2})$, then

$$\text{Sup } R_{\hat{\mu}} = 1+u \qquad \text{- - - - - (24)}$$

where $u > 0$ is a quantity whose order of magnitude is larger than $n^{-1/3}$. On the other hand, it follows from (17) that for $m = O(n^{2/3})$

$$\text{Sup } R_{\hat{\mu}} = 1 + O(n^{-1/3}) \qquad \text{- - - - - (25)}$$

We conclude that $m = O(n^{2/3})$ is the asymptotic minimax value of m for which

$$\lim_{n \to \infty} \text{Sup } R_{\hat{\mu}} = 1 \qquad \text{- - - - - (26)}$$

Also, it is clear from (22) that for any m a necessary condition for (26) to hold is that

$$m \to \infty \text{ and } \frac{m}{n} \to 0 \quad \text{as} \quad n \to \infty \qquad \text{- - - - - (27)}$$

6. Supremum of $R_{\hat{\mu}}$. Let I be the interval defined as

$$\theta_o \leq \theta_i \leq 1 \qquad i=1,\ldots,k \qquad ----- (28)$$

where $\theta_o = 1 - \frac{2t_o}{\sqrt{m}}$ and $t_o = O(1)$. It is shown below that $R_{\hat{\mu}}$ is maximized either inside the interval I or at $\psi = \psi_o = (o,\ldots,0,1)$.

We have

$$R_1(\psi) = E \frac{n}{\sum_{i=1}^{k} n_i\theta_i}$$

$$= \sum_{i=1}^{k} \frac{n}{m \sum_{\alpha=1}^{k} \theta_\alpha + (n-mk)\theta_i} p\{n_i = n-mk+m\}$$

$$\leq \frac{n}{n-mk+m} \sum_{i=1}^{k} \frac{1}{\theta_i} p\{n_i = n-mk+m\}$$

$$\leq \frac{n}{n-mk+m} \sum_{i=1}^{k} \frac{1}{\theta_i} p\{F \leq \theta_i\} \qquad ----- (29)$$

The last inequality follows from (13). The density function of F, which we denote by g(F), is unimodal with its mode at $F = \frac{m-3}{m+1}$. Hence, by the mean value theorem, for $0 \leq \theta \leq \frac{m-3}{m+1}$

$$p\{F \leq \theta\} \leq \theta g(\theta)$$

From (29) we get, therefore,

$$R_1(\psi) \leq \frac{n}{n-mk+m} \{\tfrac{1}{2} + \sum_{i=1}^{k-1} g(\theta_i)\}$$

$$\leq \frac{n}{n-mk+m} \{\tfrac{1}{2} + (k-1)\, g(\theta_o)\} \qquad ----- (30)$$

if $\theta_i \leq \theta_o \leq \frac{m-3}{m+1}$ $i=1,\ldots,k-1$. By Edgeworth's approximation of density function ([2] page 265) we get

$$g(\theta_o) = g(1 - \frac{2t_o}{\sqrt{m}}) \doteq \varphi(-t_o) + O(\frac{1}{\sqrt{m}})$$

where $\varphi(t) = \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}$. Hence when m is sufficiently large we can choose the value of t_o such that $(k-1)g(\theta_o) \leq \frac{1}{2}$. From (30) it follows that, when m is sufficiently large and t_o is suitably determined then

$$R_1(\psi) \leq \frac{n}{n-mk+m}$$

if $\psi \notin I$. But from (18) we have

$$R_{\hat{\mu}}-R_1 = O(\frac{1}{n})$$

Hence, when m is sufficiently large and $\psi \notin I$

$$R_{\hat{\mu}}(\psi) \leq \frac{n}{n-mk+m}$$

But

$$R_{\hat{\mu}}(\psi_o) = \frac{n}{n-mk+m}$$

Therefore, we conclude that

<u>Theorem 3</u> $\quad \operatorname{Sup} R_{\hat{\mu}}(\psi) = \max\{R_{\hat{\mu}}(\psi_o), \sup_{\psi \in I} R_{\hat{\mu}}(\psi)\}$

For k=2 we obtain the above result more directly. For $\theta \leq 1$ we have from (6)

$$R_\mu = n\{\int_0^{1/\theta} \frac{n-m+mF^2\theta}{(n-m+mF\theta)^2}\, dG(F) + \int_{1/\theta}^{\infty} \frac{m+(n-m)F^2\theta}{(m+(n-m)F\theta)^2}\, dG(F)\}$$

where G(F) denotes the cumulative distribution function of F.

Differentiating with respect to θ we get

$$\frac{\partial R_{\hat{\mu}}}{\partial \theta} = \frac{(n-2m)(1-\theta)}{n\theta^3} g(\theta) + nE_F \frac{n_1 n_2 F(F-2) - n_2^2 F^3 \theta}{(n_1 + n_2 F\theta)^3} \quad \text{- - - - - (31)}$$

$$\frac{\partial^2 R_{\hat{\mu}}}{\partial \theta^2} = \frac{(n-2m)g(\theta)}{2n\theta^4(1+\theta)} \{(m+5)\theta^2 - 2(m+1)\theta + m-11\}$$

$$+ nE_F \frac{2n_2^3 F^4 \theta - 2n_1 n_2^2 F^2(2F-3)}{(n_1 + n_2 F\theta)^4} \quad \text{- - - - - (32)}$$

where $g(F)$ denotes the density function of F.

Now, for $F \leq 1/\theta$

$$n_1 + n_2 F\theta = n-m+mF\theta \leq n \ ,$$

for $F > 1/\theta$

$$n_1 + n_2 F\theta = m+(n-m)F\theta > n$$

Hence for any θ, $n_1 + n_2 F\theta$ is an increasing function of F. By the mean value theorem

$$E_F \frac{F(F-2)}{(n_1 + n_2 F\theta)^3} = \frac{1}{(n-m)^3} \int_0^{\xi} F(F-2)dG(F)$$

where ξ is some positive number. Now

$$F(F-2) \lesseqqgtr 0 \quad \text{according as } F \lesseqqgtr 2$$

and

$$\int_0^{\infty} F(F-2)dG(F) = -\frac{(m-1)(m-11)}{(m-3)(m-5)} < 0 \qquad \text{for } m > 11$$

$$\therefore \int_0^{\xi} F(F-2)dG(F) < 0 \qquad \text{for all positive } \xi, \qquad m > 11$$

that is,

$$E_F \frac{F(F-2)}{(n_1 + n_2F\theta)^3} < 0 \qquad \text{if } m > 11$$

Since the first term on the right of (31) is equal to zero for $\theta=0$ and $m > 11$, it follows that $R_{\hat{\mu}}$ is a decreasing function of θ at $\theta=0$. Again

$$E_F\{F^2(2F-3)\} = -\frac{(m+1)(m-1)(m-27)}{(m-3)(m-5)(m-7)} < 0 \qquad \text{for } m > 27$$

Therefore, by similar argument as given above

$$E_F \frac{n_2F^2(2F-3)}{(n_1 + n_2F\theta)^4} < 0 \qquad m > 27$$

that is, the second term on the right of (32) is always positive if $m > 27$. Let

$$\theta = 1 - \frac{2t}{\sqrt{m}} \quad ;$$

the quantity inside the brace of the first term on the right of (32) is equal to

$$\frac{4(m+5)}{m} t^2 - 8 - \frac{16t}{\sqrt{m}} > 4t^2 - 16 \geq 0 \qquad \text{for } t \geq 2$$

Hence

$$\frac{\partial^2 R_{\hat{\mu}}}{\partial \theta^2} > 0 \qquad \text{for } 0 \leq \theta \leq 1 - \frac{4}{\sqrt{m}} \qquad \text{for } m > 27$$

Thus we have shown that $R_{\hat{\mu}}$ is decreasing at $\theta=0$ and is convex inside the interval

$$0 \leq \theta \leq 1 - \frac{4}{\sqrt{m}}$$

Hence, $R_{\hat{\mu}}$ is maximized either inside this interval or at $\theta=0$.

7. <u>The Minimax Value of m.</u> It was shown in section 5 that $m = O(n^{2/3})$ is asymptotic minimax value of m. By theorem 3, $R_{\hat{\mu}}(\psi)$ is maximized either at $\psi=\psi_o$ for which its value is $\frac{n}{n-mk+m}$, or somewhere inside the interval I given by (28). Therefore, the minimax value of m is a solution of the equation

$$\frac{n}{n-mk+m} = \sup_{\psi \in I} R_{\hat{\mu}}(\psi) \qquad \text{- - - - - (33)}$$

By theorem 1, $R_{\hat{\mu}}(\psi)$ is a symmetric function of the arguments $\theta_1,\ldots,\theta_k$. Also, when m is large, I is a small interval. Hence in looking for the supremum of $R_{\hat{\mu}}(\psi)$ inside I we restrict ψ to those points in I which $\theta_1 = \theta_2 = \ldots = \theta_{k-1} = \theta$, say. We shall denote such a point by ψ'. Then an asymptotic minimax value of m is a solution of the equation

$$\frac{n}{n-mk+m} = \sup_{\psi' \in I} R_{\hat{\mu}}(\psi') \qquad \text{- - - - - (34)}$$

To evaluate the quantity on the right of (33) we first compute $\sup_{\psi' \in I} R_1(\psi')$. From (9) we get

$$R_1(\psi') = \frac{n}{n-mk+m+m(k-1)\theta} \int_0^\infty \{1-v(x\theta)\}^{k-1} dv(x)$$

$$+ \frac{n}{m+(n-m)\theta} [1 - \int_0^\infty \{1-v(x\theta)\}^{k-1} dv(x)] \qquad \text{- - - (35)}$$

$$= 1 + \frac{n(1-\theta)}{m+(n-m)\theta} \left[\frac{n-m}{n} - \frac{(n-mk)}{n-mk+m+m(k-1)\theta} \int_0^\infty \{1-v(x\theta)\}^{k-1} dv(x) \right] \qquad \text{- - - - (36)}$$

where $v(x)$ denotes the cumulative distribution function of chi-square variable with m-1 degrees of freedom. Put $\theta = 1 - \frac{2t}{\sqrt{m}}$ where $t = 0(1)$. From (36) we have then, using Edgeworth's asymptotic approximation of distribution functions

$$R_1(\psi') = 1 + \frac{2t}{\sqrt{m}} \left\{ \frac{n-m}{n} - \frac{n-mk}{n} \int_{-\infty}^{\infty} \Phi^{k-1}(-u+\sqrt{2}\, t)\, \varphi(u)du + 0(\tfrac{1}{m}) \right\}$$

where $\Phi(x) = \int_{-\infty}^{x} \varphi(x)dx$. Let

$$c_k = \operatorname*{Sup}_t 2t\{1 - \int_{-\infty}^{\infty} \Phi^{k-1}(-u+\sqrt{2}\, t)\, \varphi(u)du\} \qquad \text{- - - - - (37)}$$

Then,

$$\operatorname*{Sup}_{\psi' \in I} R_1(\psi') = 1 + \frac{c_k}{\sqrt{m}} + 0(\tfrac{1}{m}) \qquad \text{- - - - - (38)}$$

It has been shown earlier that $R_{\hat{\mu}} - R_1 = 0(\frac{1}{n})$. We have then from (38)

$$\operatorname*{Sup}_{\psi' \in I} R_{\hat{\mu}}(\psi') = 1 + \frac{c_k}{\sqrt{m}} + 0(\tfrac{1}{m}) \qquad \text{- - - - - (39)}$$

From (34) and (39) we find, neglecting terms of order $\frac{1}{m}$ and higher, that the asymptotic minimax value of m, which we shall denote by m_k, is a solution of the equation

$$\frac{n}{n-mk+m} = 1 + \frac{c_k}{\sqrt{m}} \qquad \text{- - - - - (40)}$$

This gives

$$m_k = \left(\frac{c_k n}{k-1}\right)^{2/3} + o(n^{1/3}) \qquad \text{- - - - - (41)}$$

Also, for the corresponding value of the minimax risk we have

$$\underset{m}{\text{Inf}}\ \underset{\psi'}{\text{Sup}}\ R_{\hat{\mu}}(\psi') = \frac{n}{n-m_k(k-1)} \qquad \text{- - - - - (42)}$$

For the case of k=2 populations we have carried out the approximation one stage further. Neglecting terms of order smaller than $\frac{1}{m}$ we find that the minimax value of m is a solution of the equation

$$\frac{n}{n-m} = 1 + \underset{t}{\text{Sup}}\left[\frac{2t\,\Phi(-t)}{\sqrt{m}} + \frac{2t^2}{m}\{2\Phi(-t) - t\varphi(t)\}\right] \qquad \text{- - - - - (43)}$$

Let $c_2 = \underset{t}{\text{Sup}}\ 2t\,\Phi(-t) = 2t_2\,\Phi(-t_2)$. Then we approximate (43) by

$$\frac{n}{n-m} = 1 + \frac{c_2}{\sqrt{m}} + \frac{2t_2^2}{m}\{2\Phi(-t_2) - t_2\varphi(t_2)\}$$

$$= 1 + \frac{c_2}{\sqrt{m}} + \frac{d}{m} \qquad \text{- - - - - (44)}$$

Approximately, $c_2 = .3399$ $\quad d = .2556$ $\quad t_2 = .7518$

To summarize we have shown that

Theorem 4 The minimax value of m, denoted by m_k, is a solution of the equation

$$\frac{n}{n-mk+m} = \underset{\psi\,\in\,I}{\text{Sup}}\ \mathcal{R}$$

which gives $m_k = \left(\frac{c_k n}{k-1}\right)^{2/3} + o(n^{1/3})$, where c_k is a constant. $c_2 = .3399$ approximately. The minimax risk is equal to $\frac{n}{n-m_k(k-1)}$.

8. Single-Stage Sampling Rule. In single-stage sampling procedure with a fixed total of n available observations, we take a sample of size $\frac{n}{k}$ from each population, assuming that n is a multiple of k. The equality of the sample size is required by the invariance of the sample size funtion under permutations of the labels of the populations. By the explanation given earlier in section 3, this is a minimax rule.

Analogous to (2) an estimator of μ, which we shall denote by $\hat{\mu}_o$ is defined as

$$\hat{\mu}_o = \frac{\sum_{i=1}^{k} \frac{n_i \bar{x}_i}{s_i^2}}{\sum_{i=1}^{k} \frac{n_i}{s_i^2}} = \frac{\sum_{i=1}^{k} \frac{\bar{x}_i}{s_i^2}}{\sum_{i=1}^{k} \frac{1}{s_i^2}} \qquad \text{- - - - - (45)}$$

where $n_i = \frac{n}{k}$; $\bar{x}_i$ and s_i^2 are the mean and variance, respectively, of the sample from π_i $(i=1,\ldots,k)$. $\bar{x}_i$ and s_i^2 are based on $\frac{n}{k}$ observations for each i. It is shown below that the supremum of the risk of $\hat{\mu}_o$, which we shall denote by $R_{\hat{\mu}_o}$, is equal to k. This compares unfavorably with the supremum of $R_{\hat{\mu}}$ which, as we have seen in the previous section, is asymptotically equal to 1. The difference between the two values shows the advantage of the two-stage sampling procedure over the single-stage sampling rule under the criterion of minimaxity.

We easily compute

$$R_{\hat{\mu}_o} = k\, E \frac{\sum_{i=1}^{k} \frac{\theta_i}{x_i^2}}{\left(\sum_{i=1}^{k} \frac{\theta_i}{x_i} \right)^2} \qquad \text{- - - - - (46)}$$

where $x_1,\ldots,x_k$ are k independent chi-square variables, each of them having $\frac{n}{k} - 1$ degrees of freedom. Differentiating partially with respect to θ_i we have

$$\frac{1}{k}\frac{\partial R_{\hat{\mu}_o}}{\partial\theta_i} = E\frac{\sum_{j\neq i}\theta_j\left(\frac{1}{x_i}-\frac{2}{x_j}\right)\frac{1}{x_j x_i}-\frac{\theta_i}{x_i^3}}{\left(\sum_{i=1}^{k}\frac{\theta_i}{x_i}\right)^3} \quad \text{- - - - - (47)}$$

Consider a particular term of the summation on the right of (47). Let

$$A = E\frac{\theta_j\left(\frac{1}{x_i}-\frac{2}{x_j}\right)\frac{1}{x_j x_i}}{\left(\sum_{i=1}^{k}\frac{\theta_i}{x_i}\right)^3}$$

$$= E\frac{\theta_j\frac{x_j}{x_i}\left(\frac{x_j}{x_i}-2\right)}{\left\{\theta_i\frac{x_j}{x_i}+\theta_j+x_j\left(\sum_{h\neq i,j}\frac{\theta_h}{x_h}\right)\right\}^3}$$

$$\leqq E\frac{\theta_j\frac{x_j}{x_i}\left(\frac{x_j}{x_i}-2\right)}{\left\{2\theta_i+\theta_j+x_j\left(\sum_{h\neq i,j}\frac{\theta_h}{x_h}\right)\right\}^3}$$

$$= E\frac{\frac{\theta_j x_j}{\frac{n}{k}-3}\left(\frac{x_j}{\frac{n}{k}-5}-2\right)}{\left\{2\theta_i+\theta_j+x_j\left(\sum_{h\neq i,j}\frac{\theta_h}{x_h}\right)\right\}^3}$$

Note that $E\frac{1}{x_i}=\frac{1}{\frac{n}{k}-3}$; $E\frac{1}{x_i^2}=\frac{1}{\left(\frac{n}{k}-3\right)\left(\frac{n}{k}-5\right)}$. A second step gives

$$A \leqq E\frac{\frac{k\,\theta_j x_j}{n-3k}\left(\frac{k\,x_j}{n-5k}-2\right)}{\left\{2\theta_i+\theta_j-\frac{2(n-5k)}{k}\sum_{h\neq i,j}\frac{\theta_h}{x_h}\right\}^3}$$

$$= \frac{\frac{(n-k)\theta_j}{n-3k}\left(\frac{n+k}{n-5k} - 2\right)}{\{2\theta_i + \theta_j - \frac{2(n-5k)}{k} \sum_{h \neq i,j} \frac{\theta_h}{k_h}\}^3}$$

$$< 0 \quad \text{if} \quad n > 11k$$

From (47) it follows that $R_{\hat{\mu}_o}(\theta_1, \ldots, \theta_k)$ is decreasing in each θ_i ($i \neq k$) provided $n > 11k$. Hence putting $\theta_1 = \ldots = \theta_{k-1} = 0$ in (46) we get

$$\text{Sup } R_{\hat{\mu}_o} = R_{\hat{\mu}_o}(0, \ldots, 0, 1) = k \qquad \text{- - - - - (48)}$$

Similarly, putting $\theta_1 = \ldots = \theta_k = 1$ we get

$$\text{Inf } R_{\hat{\mu}_o} = R_{\hat{\mu}_o}(1, \ldots, 1) = k\, E \frac{\sum_{i=1}^{k} \frac{1}{x_i^2}}{\left(\sum_{i=1}^{k} \frac{1}{x_i}\right)^2} \qquad \text{- - - - - (49)}$$

For k=2 we have from (49)

$$\text{Inf } R_{\hat{\mu}_o} = 2E \frac{1+F^2}{(1+F)^2} = \frac{n+2}{n} \qquad \text{- - - - - (50)}$$

9. Use of the Second Sample Variance. It will be observed from (2) that the quantities s_i^2, $(i=1, \ldots, k)$ obtaining in the expression for $\hat{\mu}$ are used as estimates of the variances of the k populations. These estimates are obtained from the results of the first-stage experiment. We might use the second sample variance to improve the estimate of σ^2, the smallest of the population variances, especially when the second sample is large. Let $\hat{\delta}$ be obtained by substituting the second sample variance for the smallest of $\bar{s}_1^2, \ldots, s_k^2$ in $\hat{\mu}$. We should expect that the risk of $\hat{\delta}$, which shall denote by $R_{\hat{\delta}}$ is smaller than $R_{\hat{\mu}}$ when m is small compared to n-mk. We shall show that this is true for $\psi = (1, \ldots, 1)$. On the other hand,

it is easy to see that for $\psi=\psi_o=(0,\ldots,0,1)$

$$R_{\hat{\mu}}(\psi_o) = R_{\hat{\delta}}(\psi_o) = \frac{n}{n-mk+m}$$

Also, both $R_{\hat{\mu}}$ and $R_{\hat{\delta}}$ approach R_1 when m and n-mk become large.

Let ψ' denote the point $(1,\ldots,1)$. We easily compute

$$\frac{1}{n} R_{\hat{\mu}}(\psi') = k\, E\left\{ \frac{m \sum\limits_{\alpha \neq i} \frac{1}{x_\alpha^2} + \frac{n-mk+m}{x_i^2}}{\left(m \sum\limits_{\alpha \neq i} \frac{1}{x_\alpha} + \frac{n-mk+m}{x_i}\right)^2} \,\middle|\, n_i = n-mk+m\right\} p\{n_i=n-mk+m\}$$

$$= \frac{1}{n-mk+m} E\left\{1+ \frac{m(k-1)}{n-mk+m}\left(\frac{x_i^2}{x_\alpha^2} - \frac{2x_i}{x_\alpha}\right) \,\middle|\, n_i=\min(x_1,\ldots,x_k)\right\}$$

$$+ O\left(\frac{m^2}{(n-mk+m)^3}\right) \qquad \text{- - - - - (51)}$$

where $x_1,\ldots,x_k$ are k independent chi-square variables each with m-1 degrees of freedom and x_i denotes the smallest of them. When k is moderately large, $\frac{x_i}{x_\alpha}$ is comparable to $F_k = \frac{x_i}{x}$ where x is another chi-square variable with m-1 degrees of freedom and independent of $x_1,\ldots,x_k$. The moments and percentile points of F_k have been tabulated by S. Gupta and M. Sobel ([8] pages 509-523), from which we calculate $E(F_k^2-2F_k)$ for several values of m and k. These are shown below:

$-E(F_k^2-2F_k)$

k =	2	3	4	5	6
m = 9	.25	.35	.53	.55	.58
= 11	.51	.62	.66	.67	.67

Let $-c_{m,k} = E\{(\frac{x_i^2}{x_\alpha^2} - \frac{2x_i}{x_\alpha})|n_i=\min(x_1,\ldots,x_k)\}$. Then the values of $c_{m,k}$ are approximately those shown in the above table. It will be seen that $c_{m,k}$ is appreciably less than 1. From (51) we have

$$\frac{1}{n} R_{\hat{\mu}}(\psi') = \frac{1}{n-mk+m} \{1 - \frac{m(k-1)\ c_{m,k}}{n-mk+m}\} + O(\frac{m^2}{(n-mk+m)^3}) \quad ------ (52)$$

For the risk of $\hat{\delta}$ we have

$$\frac{1}{n} R_{\hat{\delta}}(\psi') = k\ E\{ \frac{m \sum\limits_{\alpha \neq i} \frac{1}{x_\alpha^2} + \frac{n-mk+m}{x_o^2}}{(m \sum\limits_{\alpha \neq i} \frac{1}{x_\alpha} + \frac{n-mk+m}{x_o})^2} | n_i=n-mk+m\} p\{n_i=n-mk+m\}$$

$$+ O(\frac{m^2}{(n-mk+m)^3})$$

$$= \frac{1}{n-mk+m} E\{1+ \frac{m(k-1)}{n-mk+m} (\frac{x_o^2}{x_\alpha^2} - \frac{2x_o}{x_\alpha})|n_i=\min(x_1,\ldots,x_k)\}$$

$$+ O(\frac{m^2}{(n-mk+m)^3}) \quad ----- (53)$$

where $x_o = \frac{mx'}{n-mk-1}$ where x' is a chi-square variable with n-mk-1 degrees of freedom and independent of $(x_1,\ldots,x_k)$. When k is moderately large, $\frac{x_o}{x_\alpha}$ $(\alpha \neq i)$ is, approximately, distributed as $F_{n-mk-1,\ m-1}$ the weighted ratio of two independent chi-square variables with n-mk-1 and m-1 degrees of freedom. Hence approximately,

$$E\{(\frac{x_o^2}{x_\alpha^2} - \frac{2x_o}{x_\alpha})|n_i=\min(x_1,\ldots,x_k)\}$$

$$= E(F^2_{n-mk-1,\ m-1} - 2F_{n-mk-1,\ m-1})$$

$$= -1 - \frac{2}{m-5} + O(\frac{1}{n-mk-1})$$

Hence from (53) we have

$$\frac{1}{n} R_{\hat{\delta}}(\psi') = \frac{1}{n-mk+m} \{1 - \frac{m(k-1)}{n-mk+m} (1 + \frac{2}{m-5})\} + 0(\frac{m^2}{(n-mk+m)^3}) \text{- - -} (54)$$

From (52) and (54) we find

$$R_{\hat{\mu}}(\psi') \geqq R_{\hat{\delta}}(\psi')$$

the desired result.

10. Uniform Populations. In this section we consider uniform populations with a common mean. The formulation of the problem is the same as in the case of normal populations which we have discussed earlier. The analysis is similar. We shall, therefore, give some of the main results without showing the derivation.

Let $\pi_1,\ldots,\pi_k$ be k rectangular populations with a common mean equal to μ. We shall denote the range of π_i by a_i $(i=1,\ldots,k)$. The quantities a_i and μ are unknown. It is required to estimate μ by a two-stage sampling procedure. The total number of available observations is fixed, equal to n, say. We take $m \geqq 2$ observations from each population in the first-stage experiment, compute the sample range for each population and take n-mk observations in the second-stage experiment from the population corresponding to the smallest range of the first sample. The estimate of μ is then given by

$$\hat{\mu} = \frac{x_{(1)} + x_{(n-mk)}}{2}$$

where $x_{(1)}$, $x_{(n-mk)}$ are the extreme values of the observations in the second sample. The risk of $\hat{\mu}$, which we shall denote by R, is defined as

$$R = \frac{2(n+1)(n+2)}{a^2} E(\hat{\mu}-\mu)^2 \qquad \text{- - - - -} (56)$$

where $a = \min(a_1,\ldots,a_k)$.

Let $x_{i(1)}$, $x_{i(m)}$ be the smallest and largest values, respectively, in a sample of m observations from π_i. Then the pair $x_{i(1)}$, $x_{i(m)}$ is a sufficient statistic for $(\mu,\ a_i)$ whose estimates are given by

$$\hat{\mu}_i = \frac{x_{i(1)} + x_{i(2)}}{2}$$

$$r_i = x_{i(m)} - x_{i(1)}$$

It is easy to show that

$$E(\hat{\mu}_i) = \mu$$

$$v(\hat{\mu}_i) = \frac{a_i^2}{2(m+1)(m+2)}$$

and that the density and cumulative distribution function of r_i are given by

$$f(r_i=y) = \frac{m(m-1)}{a_i^m}(a_i-y)\,y^{m-2} \qquad 0<y \leqq a_i$$

$$= 0 \qquad y > a_i$$

$$F(r_i=y) = \frac{m(m-1)}{a_i^m}\left(\frac{a_i}{m-1} - \frac{y}{m}\right)y^{m-1} \qquad 0<y \leqq a_i$$

$$= 1 \qquad y > a_i$$

From (56) we compute

$$\frac{(n-mk+1)(n-mk+2)R}{(n+1)(n+2)m(m-1)}$$

$$= \sum_{i=1}^{k} \theta_i^2 \int_0^{1/\theta_i} \prod_{j\neq i} \{1-(m-(m-1)\frac{\theta_i x}{\theta_j})(\frac{\theta_i x}{\theta_j})^{m-1}\} x^{m-2}(1-x)dx$$

where $\theta_i = \frac{a_i}{a}$ $(i=1,\ldots,k)$. Notice that $\theta_i \geq 1$ and that equality is satisfied for at least one value of i. We shall now indicate some of the main results.

(i) A necessary condition for $\lim_{n \to \infty} \text{Sup } R = 1$ is that

$$m \to \infty \quad \text{and} \quad \frac{m}{n} \to 0 \qquad \text{as } n \to \infty$$

(ii) $R(\psi)$ is a symmetric function of the arguments $\theta_1,\ldots,\theta_k$.

(iii) $R(\psi)$ is maximized inside the interval I' defined as

$$1 \leq \theta_i \leq 1+h \qquad i=1,\ldots,k$$

where $h = O(\frac{1}{m})$. Let ψ' be the maximizing value of ψ. We assume that $\psi = (\theta,\ldots,\theta,1)$ for some value of θ. Then

$$\frac{(n-mk+1)(n-mk+2)R(\psi')}{(n+1)(n+2)m(m-1)} - \frac{1}{m(m-1)}$$

$$= (\theta_k^2-1)\theta_k^{m-1} \sum_{r=1}^{k-1} (-1)^{r+1} \binom{k-1}{r} \sum_{S=0}^{r} (-1)^S m^{r-S} (m-1) \binom{r}{S} \frac{\theta_k^{-(m-1)(r+1)-S}}{((m-1)(r+1)+S)((m-1)(r+1)+S+1)}$$

For an asymptotic solution of θ we have $\theta = 1 + \frac{h_k}{m}$ where h_k is the value of h maximizing

$$\frac{1}{m(m-1)^2} \sum_{r=1}^{k-1} \frac{2(-1)^{r+1} h e^{-hr} (1+h)^r \binom{k-1}{r}}{(r+1)^2} \{1 + \frac{2r}{(1+h)(r+1)}\}$$

We shall denote the maximum value by u_k.

(iv) $$\text{Sup } R = \frac{(n+1)(n+2)(1 + \frac{u_k}{m})}{(n-mk+1)(n-mk+2)} + O(\frac{1}{m^2})$$

(v) The asymptotic minimax value of m, which we denote by m_k, is given by

$$m_k = \left(\frac{nu_k}{2k} \right)^{\frac{1}{2}} + O(n^{\frac{1}{4}})$$

and

$$\operatorname*{Inf}_m \operatorname{Sup} R = \frac{(n+1)(n+2)(1 + \frac{u_k}{m_k})}{(n-m_k k+1)(n-m_k k+2)}$$

(vi) For k=2

$$\theta = 1 + \frac{\sqrt{2}}{m} + \frac{2+\sqrt{2}}{m^2} + O\left(\frac{1}{m^3} \right)$$

$$u_2 = (1 + \sqrt{2})\, e^{-\sqrt{2}} = .5871 \text{ approximately}$$

Corresponding to one-stage sampling procedure the maximum value of the risk is

$$\frac{13n^2-10n-8}{2(n-1)(n-4)} \left\{1 + \frac{52}{n^2} + O\left(\frac{1}{n^3} \right)\right\}$$

(vii) Compare the above results with those obtained for the normal populations. Here the minimax value of m is of order $n^{1/2}$ as against $n^{2/3}$ in the normal case. Also, the minimax risk is equal to $1 + O\left(\frac{1}{\sqrt{n}} \right)$ as against $1 + O\left(\frac{1}{n^{1/3}} \right)$.

Acknowledgment. The author is deeply indebted to Professor M. Sobel and Professor I. R. Savage for their valuable advice in the course of writing this paper.

REFERENCES

[1] Blackwell, David and Girshick, M. A. (1954). Theory of Games and Statistical Decisions, John Wiley and Sons, New York.

[2] Wilks, S. S. (1962). Mathematical Statistics, John Wiley and Sons, New York.

[3] Richter, D. (1960). "Two-stage experiments for estimating a common mean." Ann. Math. Statist. 31 pp. 1164-1173.

[4] Ghurye, S. G. and Robbins, H. (1954). "Two-stage procedures for estimating the difference between means." Biometrika 41 pp. 146-152.

[5] Wald, A. (1950). Statistical Decision Functions, John Wiley and Sons, New York.

[6] Hunt, G. and Stein, C. "Most stringent tests of composite hypotheses", Unpublished Paper.

[7] Lehman, E. L. (1950). Lecture--"Notes on the theory of estimation", University of California.

[8] Gupta, S. and Sobel, M. (1962). "On the smallest of several correlated F statistics", Biometrika 49 pp. 509-523.

APPENDIX

Values of c_k defined in (37) and used in (41).

k	c_k
2	.340
3	.527
4	.665
5	.752